DOCUMENT RESUME

ED 153 470 FL 009 386

AUTHOR Lowe, Pardee, Jr.
TITLE The Oral Language Proficiency Test.
PUB DATE Dec 76
NOTE 38p.; Accompanying tapes are available from the National Audiovisual Center; Prepared as part of the Interagency Language Roundtable; Contains occasional small print

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Audiolingual Skills; Communicative Competence (Languages); Definitions; *Interviews; *Language Instruction; *Language Proficiency; *Language Tests; Listening Comprehension; Measurement Techniques; Questioning Techniques; *Second Language Learning; Speech Communication; Speech Skills; Testing; Video Tape Recordings
IDENTIFIERS *Oral Proficiency Test

ABSTRACT

This booklet describes the test used by the federal government to determine oral proficiency in a foreign language. The test is essentially an oral interview, conducted by two testers, and lasting from 15 to 30 minutes, although the format can vary. Proficiency levels range from 0 (for no practical speaking proficiency) to 5 (for speaking proficiency equivalent to that of an educated native speaker). Definitions of rating levels are stringent, and testers are carefully trained in testing methods and proficiency definitions. Accompanying videotape cassettes demonstrate how the test should and should not be administered. Commentaries on both are appended, as well as a detailed description of the FSI (Foreign Service Institute) Checklist of Performance Factors. (AM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED153470

THE ORAL LANGUAGE PROFICIENCY TEST

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Pardee Lowe, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

by

Pardee Lowe, Jr.

INTERAGENCY LANGUAGE ROUNDTABLE

U.S. GOVERNMENT

Washington, D.C.
December 1976

ED009386

Acknowledgements

I wish to express my appreciation to all the participating agencies in the Interagency Language Roundtable who so freely shared their knowledge and experience in oral interview testing, particularly to Marianne Adams, Dorothy Rauscher, and Claudia P. Wilds, and to our testing colleagues at ETS, John L. D. Clark, Russell Webster, and Protase E. Woodford who viewed an earlier version of the videocassette and also shared their views on the accompanying booklet. Any remaining errors of fact or interpretation are solely the responsibility of the author.

For further information on the oral interview testing technique, write:

Testing Committee
INTERAGENCY LANGUAGE ROUNDTABLE
Box 9212
Rosslyn Station, Virginia 2[illegible]9

CONTENTS

	Page
I. Introduction to Oral Proficiency Testing	1
II. Format of the Oral Interview	2
A. General	2
B. Two-Phase Version	3
C. Four-Phase Version	4
D. Intermediate Versions	5
III. Proficiency Definitions	5
A. General	5
B. Misunderstandings	7
C. Reliability of the Test	10
IV. Rating the Candidate	10
A. FSI Rating	10
B. LLC Rating	11
V. Testers	13
VI. Elicitation Technique	14
VII. Videocassette	15
VIII. Commentaries on the Taped Tests	16
Appendix: Detailed Description of the FSI Checklist	29
Bibliography on the Oral Interview	35

I. Introduction to Oral Proficiency Testing

The Second World War generated an intense interest in spoken foreign language in the United States and led to a marked shift in high school and college foreign language programs from emphasis on Reading and Writing to emphasis on Speaking and Understanding. New teaching materials were designed to accommodate the change and were adopted by many foreign language programs outside the government. Despite this shift, no adequate testing instrument for spoken foreign language competence has emerged in academia. Most tests currently in use are achievement tests, often based on specific texts. Generally, these tests do not test speaking competence directly. There are a few standardized proficiency tests which contain a separate section for evaluating speaking competence.[1] However, they do not test speaking as extensively as the test described in this booklet. Where businessmen claim foreign language speaking competence, their firms most often rely on a man's own estimation of his ability. No valid instrument to assess functional skills in the language is available.

Facing the same problem, the U.S. Department of State Foreign Service Institute devised in the mid-1950s a set of Proficiency Definitions stressing functional use of spoken foreign language. It then developed a testing program to rate the language skills of Foreign Service Officers and other government employees who use foreign languages on their jobs at home and abroad. The resulting test took the form of an oral interview,[2] a test in which two testers conduct a 15- to 30-minute conversational interview in the target language. The definitions and testing procedures are now used throughout the Federal Government, where functional speaking knowledge of a foreign langauge must be tested, primarily at the Foreign

[1] The most widely recongized proficiency test series is *The MLA Proficiency Tests for Teachers and Advanced Students*. (For a general description of the test and its availability, see Clark 1972:157.) Its Speaking Test contains three sections, two on pronunciation and one on oral production. Instructions in English ask the candidate to describe a single picture, then a series of pictures, and to react to a situation depicted pictorially. Compared to the government oral interview test described in this booklet, the MLA Speaking Test is unable to seek clarification of a candidate's previous answers, to ask for words or phrases he chose to avoid in earlier responses, to test his linguistic ability in a variety of topics and situations through actual conversation; to check for breadth of grammatical control, vocabulary, and idioms at the level at which the candidate functions best; and to push the candidate beyond his demonstrated ability to assure that his highest sustainable functional level has been correctly ascertained.

[2] No generally accepted designation for the test exists. In government it is customarily referred to as "the (oral/speaking) test." Clark (1972) calls it "The FSI Language Proficiency Interview" which is accurate, but slightly cumbersome. Jones (1975) and Wilds (1975) dub it "The Oral Interview Test." This designation suffers from a redundancy as all interviews are oral. Since all interviews have an element of testing in them, I have shortened the name to "The Oral Interview."

Service Institute (FSI), the Central Intelligence Agency (CIA), the Defense Language Institute (DLI), and the Peace Corps. The Proficiency Definitions were also accepted by the Civil Service Commission as valid for the government language community as a whole.

The present booklet and accompanying videocassette represent the first attempt to make this testing technique available to institutions outside the government. The materials illustrate how the oral interview tests spoken foreign language competence. They are a project of the Testing Committee of the Interagency Language Roundtable and are intended to acquaint new testers within the government and those outside in business and academia with the test. The booklet gives a description of the Proficiency Definitions and the oral interview, discusses elicitation technique, and provides specific commentary on the two tests on the cassette. The materials describe the common core shared by those government agencies which use the test, while also presenting some differences in format and technique. A few of the features in the videotaped tests are more characteristic of Language Learning Center (LLC) testing than that conducted at FSI. These divergencies are duly noted in the accompanying booklet. Included is a history of the present videocassette and a bibliography on the oral interview. However, we must underscore that the materials in no way replace actual training in giving tests. Government testers receive thorough training, test frequently, and recalibrate their standards when necessary. Thus, any attempt at using the oral interview in business and academia should consider such training requirements.

II. Format of the Oral Interview

A. *General*

The oral interview is a face-to-face test of foreign language speaking ability. Its shape varies with the testers and to some extent with the agency. There is a *two-phase version* current in some languages at FSI and a *four-phase version* used at the LLC.[3] Along with these two common formats, several variants exist, all of which accomplish the same task: determining where the candidate's speech sample falls on the government's scale.

The basic intention of any oral interview is to elicit from the candidate the richest possible speech sample in the shortest period of time. In a 15- to 30-minute period, two testers interview the candidate and ascertain how good his use of the target language is compared to that of an educated native

[3]Other agencies generally follow FSI's interviewing formats and elicitation practices. For clarity's sake, no attempt is made here to describe divergencies between them and FSI exhaustively. It should be noted, however, that due to limited resources, DLI generally uses only one tester who both elicits the speech sample and rates it.

speaker. The candidate's performance is rated on a scale from 0 (for no effective ability to communicate in the language) to 5 (for functioning like an educated native speaker). The scale provides +s for strong performances at all levels except Level 5, thus furnishing an 11-point scale.

In a frequently tested language like Spanish, FSI sometimes completes the test in 10 minutes with low-level (S-0 to S-1) candidates. Where the candidate shows a higher level, the test can last up to half an hour. Testers in other languages may take longer. A desirable length is 15-30 minutes; but how frequently the testers interview, how difficult a particular candidate is, and how high the level is are also contributing factors.

To assure an adequate sample in infrequently tested languages, testers extend the length of the interviews, checking points in the Definitions step-by-step. At the higher levels or in difficult interviews, tests in such languages may last 30 minutes to an hour, although this length is not encouraged.

B. *Two-Phase Version*

The two-phase version is a commonly used format in Spanish testing at FSI. It has proven to be an excellent version where tests are administered often (several per week) and all testers test constantly. From a short sample, testers can obtain a score in a minimum of time.

The test has two parts: (1) a conversation about job-related topics ("How did you come to work for the State Department? What will your overseas assignment be? How will using Spanish help on the job? What is the President's official view on a new canal through Honduras?"), and (2) a situation or series of situations where candidates role-play in the foreign language. The second part is often an interpreter situation in which the candidate carries out the instructions of a tester claiming to speak only English (the certified language examiner on the FSI team) while talking to the second tester (the native speaker at FSI) who claims to speak only the Target Language.

The first part of the test may contain relatively little probing for higher-level skills, although topics are introduced at higher levels to allow the candidate to explain American foreign policy or discuss the political situations in the country to which they will be assigned. In some instances the intensive probing at FSI comes in situations where the candidate must show whether he can handle himself in the language or not. To be maximally successful, the second part of the test must be carefully prepared ahead of time. The situations should be changed periodically so that the test is not compromised. The situations should be realistic activities that a foreigner or native speaker might actually be involved in in the target language culture.

C. *Four-Phase Version*

The four-phase interview is the standard format used at the LLC. The test begins at a low level to warm the candidate up (WARM-UP). After the warm-up, the testers begin to shift to the next higher level until they find the level at which the candidate seems to function best (LEVEL CHECK). To verify that this is in fact the highest level at which the candidate can consistently sustain a performance, the testers push the candidate to the next higher level to prove that he cannot perform at that level (PROBES). After intensive probing at the higher level, the testers return to the level where the candidate functioned best in order to leave him with a feeling of accomplishment in his performance.[4] Thus, the four-phase test moves from a WARM-UP to a LEVEL CHECK, PROBES, and finally to the WIND-UP. It is this four-phase version of testing that is illustrated in the videotaped "LLC Oral Proficiency Test." (See Table 1.)

[4]While at this point LLC and FSI test formats differ, thorough probing is a desired goal at both agencies, and neither considers a test adequate without intense questioning. Although LLC stresses a supportive ending, both agencies agree on a basic characteristic of the successful oral interview, that at the test's end all participants share the belief that a thorough and fair test has been administered and an accurate rating arrived at.

Table 1

The Four Phases of the LLC Oral Proficiency Test

LLC 1976

I. Warm-Up	II. Level Check	III. Probes	IV. Wind-Up
3-5 min The warm-up puts the candidate at ease	8-10 min The level check finds the candidate's level of speaking proficiency	7-10 min The probes find the level at which the candidate can no longer perform	2-5 min The wind-up returns the candidate to the level at which he performed best
The warm-up reacquaints the candidate with the language if necessary	The level check tests the candidate's breadth of grammatical structures, vocabulary, and idioms at that level		The wind-up leaves the candidate with a feeling of accomplishment after having been fairly and thoroughly tested
The warm-up gives the testers a preliminary indication of the candidate's level of proficiency	The level check tests for candidate's level in listening comprehension	The probes find the level at which the candidate can no longer understand	

D. *Intermediate Versions*

Both FSI and LLC have a number of intermediate versions. They generally include a Level Check and Probes, but may dispense with a Warm-Up or Wind-Up and may often substitute Situations for the Probes. The variations depend in part on the testers' sensitivity to the candidate. Since some candidates resist role-playing, situations are not always a useful technique. Testers must then substitute another testing technique which will elicit a ratable sample.

The success of the oral interview depends less on the topics discussed than on the overall elicitation technique used, i.e., the manner in which questions are posed and the types of questions asked. The same topics used over and over lose currency; they become known to the candidates through the grapevine; they eventually bore both testers and candidates. Yet, the same elicitation technique used in one Level 3 test to discuss a lack of natural resources can be reapplied in the next Level 3 interview to talk about space flight. The techniques vary more according to the candidate's level than with his personality.

However, topics can assume major importance for rating if the candidate creates answers at the same level of difficulty as the topic. For example, a question on the dangers of nuclear power by its very nature provides the opportunity for discussion at Level 3 or higher. (Of course, if the candidate answers with short, simple sentences and simplistic vocabulary, a lower level is indicated.) The central question is, "At what level does the candidate function consistently in the target language?" All topics and elicitation techniques, from the job-related conversation to the role-playing situations, are chosen to ascertain the functional level.

III. Proficiency Definitions[1]

A. *General*

The government standards are set forth in the Civil Service Proficiency Definitions below. The standards are functional and apply to all languages. Basically, they ask how well a candidate can communicate in the target language when his performance is compared to what an "educated native speaker" would say in a similar situation. "Educated" refers not to "aesthetic-literary education," but to native's ability to use the language acquired at a lycée, liceo, gymnasium, etc., in Europe or at a four-year liberal arts college or university in the USA. The interview concentrates on how the candidate expresses himself on general topics in general language, specialized topics normally being reserved for highest levels. (See Table 2.)

[1]Referred to below as "The Definitions."

Table 2

Civil Service Proficiency Definitions

Elementary Proficiency CSC 1968

S-1: Able to satisfy routine travel needs and minimum courtesy requirements. Can ask and answer questions on topics very familiar to him; within the scope of his very limited language experience can understand simple questions and statements, allowing for slowed speech, repetition, or paraphrase; speaking vocabulary inadequate to express anything but the most elementary needs; errors in pronunciation and grammar are frequent, but can be understood by a native speaker used to dealing with foreigners attempting to speak his language; while topics which are "very familiar" and elementary needs vary considerably from individual to individual, any person at the S-1 level should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, make purchases, and tell time.

Limited Working Proficiency

S-2: Able to satisfy routine social demands and limited work requirements. Can handle with confidence but not with facility most social situations, including introductions and casual conversations about current events as well as work, family, and autobiographical information; can handle limited work requirements, needing help in handling any complications or difficulties; can get the gist of most conversations on nontechnical subjects (i.e., topics which require no specialized knowledge) and has a speaking vocabularly sufficient to express himself simply with some circumlocutions; accent, though often quite faulty, is intelligible; can usually handle elementary constructions quite accurately but does not have thorough or confident control of the grammar.

Minimum Professional Proficiency

S-3: Able to speak the language with sufficient structural accuracy and vocabulary to participate in most formal and informal conversations on practical, social, and professional topics. Can discuss particular interests and special fields of competence with reasonable ease; comprehension is quite complete for a normal rate of speech; vocabularly is broad enough that he rarely has to grope for a word; accent may be obviously foreign; control of grammar good; errors never interfere with understanding and rarely disturb the native speaker.

Full Professional Proficiency

S-4: Able to use the language fluently and accurately on all levels normally pertinent to professional needs. Can understand and participate in any conversation within the range of his experience with a high degree of fluency and precision of vocabulary; would rarely be taken for a native speaker, but can respond appropriately even in unfamiliar situations; errors of pronunciation and grammar quite rare; can handle informal interpreting from and into the language.

Native or Bilingual Proficiency

S-5: Speaking proficiency equivalent to that of an educated native speaker. Has complete fluency in the language such that his speech on all levels is fully accepted by educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms, and pertinent cultural references.

B. *Misunderstandings*

The Definitions have proved meaningful to testers who understand their implications for testing technique and rating. However, they can be misleading to those reading them for the first time. The Definitions do not describe all the nuances present in the system, and only training and actual participation in interviews can provide the full framework for their interpretation. The Definitions imply a great deal more control of the target language than might be gathered at first glance. Placed for comparison in another evaluative scheme, Vallette and Disick's "Performance Objectives for Speaking" (1975:152-158), the government's Proficiency Definitions fall mainly in their *Stage 4: Communication,*[6] i.e., the oral interview demands speaking skills at the uppermost end of any communicative taxonomy.

A basic misunderstanding lies in interpreting the definitions too liberally, i.e., rating a sample of target language speech production too leniently. To deal with this problem, the Language Learning Center prepared the "LLC Guidelines for Assigning Language Proficiency Levels." The guidelines are designed to amplify the Proficiency Definitions for testers when rating. (See Table 3.)

[6]For a fuller justification for and implications of this statement, see Lowe (1975, mimeo) "The Oral Interview—A Criterion-Referenced Test?"

Table 3

For Examiners Only

LLC GUIDELINES FOR ASSIGNING LANGUAGE PROFICIENCY LEVELS

LLC 1976

Speaking - Understanding

LEVEL "0"

Speaking: The examinee has no practical speaking proficiency. He may have a few isolated words and phrases which are of no practical use.

Understanding: the examinee understands some isolated words and phrases, but is unable to participate even in a very simple conversation.

Table 3 (Continued)

LEVEL "1"

Speaking - Subject Matter. The examinee has the minimum proficiency for survival on a day-to-day basis in the target country, i.e., function in simple question-and-answer situations. He knows enough at this level to satisfy ordinary courtesy requirements. He is able to ask and answer questions relating to situations of simple daily life and routine travel abroad. The examinee is also able to request services such as renting a hotel room and ordering a simple meal. *Speaking - Quality*: the examinee at this level normally makes errors even in structures which are quite simple and common. His vocabulary is limited to the type of situations mentioned above, and even in these situations he sometimes uses the wrong word. Although his pronunciation may be poor, he makes the minimum contrastive distinctions, including stress, intonation and tone patterns, necessary to make himself understood. *Understanding*: the examinee is able to understand simple questions and statements relating to simple transactions involved in situations of daily life and independent travel abroad, allowing for slowed speech with considerable repetition or paraphrasing.

LEVEL "2"

Speaking - Subject Matter. The examinee is able to talk in some detail about concrete subjects such as his own background, his family, his job, travel experiences, recreational activities, educational background, and familiar places. *Speaking - Quality*: the examinee has enough control of the language to be able to join sentences in limited discourse. He has good control of the morphology of the language (in inflected languages), and of the most frequently used syntactical structures. Although his vocabulary is sufficient to talk with confidence about the type of topics described above, his limited vocabulary fairly often reduces him to verbal groping, or to momentary silence. A foreign intonation and rhythm may still be dominant. *Understanding*: The examinee is able to comprehend questions and statements relating to common social topics, when the language is spoken at normal conversational speed. He can get the *gist* of casual conversations with educated or well-informed native speakers talking about subjects on the level of current events, allowing for occasional repetitions or paraphrased statements.

LEVEL "3"

Speaking - Subject Matter. The examinee is able to converse and express his opinions about such topics as current events, including political and social problems of general nature. *Speaking - Quality*: The examinee has good control of grammar, though he makes occasional errors in low-frequency structures and in the most complex frequent structures. His vocabulary is broad enough so that he rarely has to grope for words in discussing the topics mentioned above. A foreign phonology, though apparent, is no longer dominant. *Understanding*: The examinee can comprehend most of what is said at a normal conversational rate of speech. A person at this level is able to understand to a high degree more complex formal discourse, i.e., subjects on the level of panel discussions, news programs, etc.

Table 3 (Continued)

LEVEL "4"

Speaking - Subject Matter: Although the subject matter that the examinee is able to handle at this level may not differ very much from that of Level "3," he is able to use the language in all nontechnical situations, and can express his opinions almost as he would in his native language (assuming that he is "5" in his native language). The examinee is able to tailor his speech to his audience. He has nearly perfect grammar and speaks the language with extensive and precise vocabulary. Although the examinee may still have an accent, he very rarely mispronounces the language. *Understanding*: the examinee can understand the content of all conversations and formal presentations within the range of his experience. With the exception of dialect variations and colloquialisms outside the range of his experience, he understands the type of language heard in speeches, sprinkled with idioms and stylistic embellishments.

LEVEL "5"

Speaking: The examinee is able to use the language in all conceivable nontechnical subjects in a manner equivalent to that of an educated or well-informed native speaker of the language. *Understanding*: The examinee is able to understand all types of formal and informal speech in a manner equivalent to that of an educated or well-informed native speaker of the language, including a wide range of dialect variations, colloquialisms, and cultural references.

Although space precludes pinpointing difficulties at each level, it is important to understand how stringent the Definitions are. For example, the Level 1 definition implies communication at the question-and-answer level for basic needs. Usually, the candidate can answer a basic question simply: a short, relatively error-free sentence. Any longer sentences will contain a larger number of errors or even be unintelligible. The candidate often can ask a simple question suitable to a basic situation, but cannot sustain a conversation, i.e., integrate the two abilities meaningfully: answering questions and asking them. In the government's experience, Level 1 already implies a level of communication beyond most introductory high school and college courses.

A second misunderstanding is that a person *must* be a "native" in the target language to be a "5." In practice, this usually proves true because most persons receiving a "5" were born in a target language country or in an enclave where a high-level version of the native language was spoken. But a small number of nonnative "5s" exist. Usually, they have lived abroad in the target language culture for many years and often speak the language at home or use it on the job.

The final misunderstanding concerns whether all native speakers of the target language are "5s." The Definitions were originally intended for rating nonnative speakers, with the speech of the "educated native speaker"

(Level 5) as the reference point. No provisions within the Definitions were made for rating native speakers who were not highly educated. However, testers are sometimes called upon to rate less educated speakers. Because the Definitions use the word "native" solely at Level 5, there is a tendency to equate automatically "native" with "5," which is demonstrably false. Thus, there are many "natives" in all languages who lack control of higher-level lexicon, syntax, and style. These speakers would never be chosen as diplomats to deal with another country at an international conference table. To rate a native with less education as "5" is to overrate him functionally (since he lacks the ability to express himself at such a high level); to rate him as a "0" is to underrate him functionally (since he clearly possesses the ability to communicate in the language). A possible solution is to rank him functionally precisely in the way nonnatives are rated, although the nature of his errors are clearly different from those of a nonnative. As yet, no uniform solution to this problem has been agreed on in the government language community.

C. *Reliability of the Test*

A discussion of misunderstandings leads to the question of the test's subjectivity. Because testers and candidates are human beings, a large number of variables play a part in the oral interview. How can they be minimized? The answer lies in carefully training testers in interviewing technique, in having them internalize the rating standards as exemplified in a calibrated set of tapes for their language, in having testers constantly discuss testing procedures and standards, and in recalibrating when the standards become vague or imprecise.

IV. Rating the Candidate

A. *FSI Rating*

At FSI, ratings are assigned in several ways. The testing team consists of a native speaker and a "certified language examiner, who may either be an experienced native-speaking language instructor or a linguist thoroughly familiar with the language (Wilds (1975):30)." The native speaker rates on his overall impression of the candidate's performance compared to that which an "educated native speaker" would produce under similar conditions. The certified language examiner, however, who is not involved in the actual questioning is often able to note the mistakes that the native interviewer might have missed. For example, the certified langauge examiner may notice that the candidate was unable to use key structures characteristic of higher levels. Although the certified language examiner

bears the final responsibility for the rating, it is arrived at in consultation with the native tester. This is called conference rating.

A more formal method is for each tester to fill out the FSI Checklist of Performance Factors given below. The candidate's performance is divided into five areas: Accent, Grammar, Vocabulary, Fluency, and Comprehension. Proportionally, the testers weight Grammar highest (3), Vocabulary and Comprehension next with (2) each, Fluency (1), and Accent (.5); and, thus, they arrive at a numerical definition of the candidate's performance.[7] (See Table 4.)

Table 4

FSI CHECKLIST OF PERFORMANCE FACTORS

FSI 1961

1. ACCENT	foreign	___	___	___	___	___	___	native
2. GRAMMAR	inaccurate	___	___	___	___	___	___	accurate
3. VOCABULARY	inadequate	___	___	___	___	___	___	adequate
4. FLUENCY	uneven	___	___	___	___	___	___	even
5. COMPREHENSION	incomplete	___	___	___	___	___	___	complete

B. *LLC Rating*

At the LLC the rating procedures are slightly different. Both testers are native speakers, and both rate on overall impression. This (as at FSI) is conditioned by on-the-job training, workshops, and reference to calibrated tapes. Scoring is done not by conference rating as at FSI, but by independently marking a rating sheet with zones representing the various levels. (See Table 5.)

The scores on the two rating sheets are then averaged by using a ruler marked in tenths of an inch. The average of the two ratings is assigned as the candidate's final rating. To prevent domination of one tester by another, testers do not discuss the candidate's performance before rating; however, discussion after the rating is encouraged. Both FSI and LLC give speaking scores. Only LLC assigns a separate score for listening comprehension.

[7] See Wilds (1975) 38-41. A fuller discussion of the Checklist of Performance Factors is to be found in the Appendix of this booklet.

Table 5

LLC RATING SHEET

LLC 1976

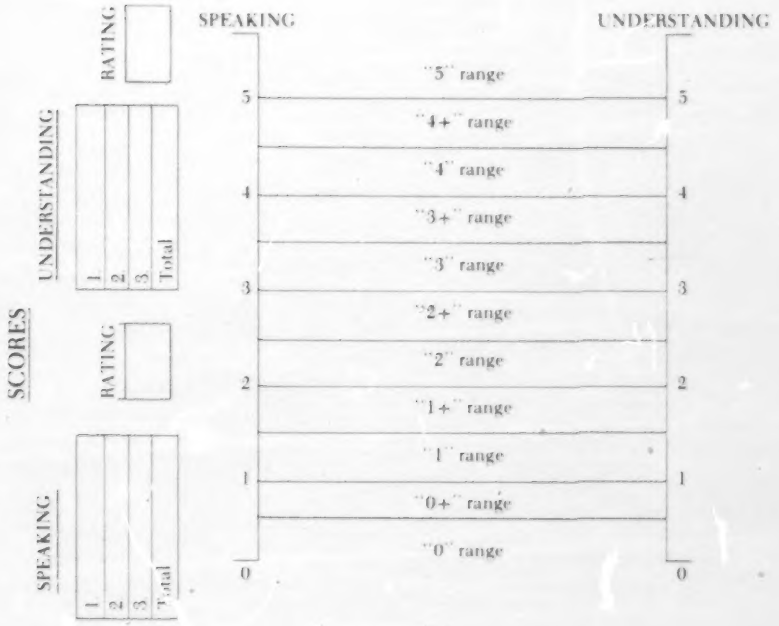
LANGUAGE PROFICIENCY RATINGS

(Oral—Aural Skills)

Examinee _____ Language _____

Examiner _____ Test Number _____

Date _____



REMARKS

Both LLC and FSI view the 11 point scale not as fixed points, but as ranges of ability. To underscore this concept the LLC Rating Sheet actually displays the ranges graphically. Thus, Level 2 is a zone where a candidate's performance can be marked as low in the range, high in the range , or in the middle of the range. The plus range is marked either for extraordinary performance, i.e., better than normal level, or for failure to sustain performance at the next highest level. At both agencies and at DLI, all tests are taped for verification by a third rater if necessary.

V. Testers

Ultimately, the success of every test rests with the interviewers. Their sensitivity to and interaction with the candidate insures the richest possible speech sample. Their sensitivity must also extend to each other's testing technique. Moreover, to conduct and rate a successful interview, they must be thoroughly familiar with the Proficiency Definitions and the rating standards for their own language.

The majority of testers are instructors of the language they test. They are either native speakers or have a thorough familiarity with the language from a linguistic point of view. Thus, even if incapable of rating on overall impression, they can produce a discrete point evaluation of the candidate's performance like a judge at a diving contest.

Just as not all native speakers of a language are good teachers, not all foreign language teachers are good oral interviewers. A few teachers instinctively understand the system. Most can acquire it through workshops and practice. Yet, there will probably be a few who will never fully learn to conduct and rate oral interviews.

A major distinction between a good tester and a good teacher is the need to interact differently with the candidate than with the student. This is one of the basic reasons why some instructors fail to succeed as oral interviewers. The helpful stance which spells success in the classroom must yield to a more objective manner in the testing room. In well-conducted tests the objective manner manifests itself in the tester's ability to let the candidate talk for himself. The psychological shift from helpfulness to objectivity is more difficult than it seems on the surface, and instructors often remind themselves consciously of the distinction before entering the testing room by rereading the Proficiency Definitions as a guide.

A good tester understands the meaning of the concept "functional use of language" in testing. The good tester knows how to conduct the test, has a good grasp of elicitation technique, can control the various question types,

and knows when and how to use them. Before the test he has prepared himself to cover a wide range of topics at the various levels. Finally, the tester understands how to rate on the 0-5 scale and has internalized the rating standards for his language. Even given this ideal profile, testers often differ. Some are stronger in one phase of testing than another. For example, a tester may have good elicitation technique, but less reliable rating. For this reason, testing in pairs is recommended. Team testing also permits one tester to observe, another to elicit.

VI. Elicitation Technique

The test procedure actually begins before the interview itself. Because most candidates are unfamiliar with the oral interview, they receive a description of the procedure before the test begins: what the interview entails, how it is conducted, how the target language is used, and how the testers will provide candidates an opportunity to demonstrate control of the target langauge. Above all, the candidates are encouraged to participate fully, to answer questions completely, and to bring up topics for discussion.

Testers should be aware at the outset that any foreign language speaker may be a candidate for the test. It is not unusual for weaker speakers and those never having participated in an oral interview to leave the testing room with a feeling of inadequacy. The basis for such a feeling is undoubtedly the test's concentration on testing oral communication and the candidate's recognition that he has failed to do so. From the very beginning of the test, interviewers should seek to minimize such effects while still probing sufficiently to acquire an adequate speech sample.

Once the test begins, one of its crucial aspects is the manner of phrasing questions in the interview. In the two tests on the videocassette, much of the second test's success rests with the way in which the testers pose their questions. Every good question has a purpose. It states a topic. It puts the topic in a broader or narrower focus. Finally, it elicits a ratable sample of speech.

Bad questions, on the other hand, waste time. They mislead the candidate. They provide the testers with no useful information for rating or repeat previously obtained material. An hour of poor questions may tell the testers little about the candidate's speech in the target langauge. Ten minutes of judicious interviewing may permit the examiner to place the candidate's speech accurately on the proficiency scale.

The manner of posing questions is equally important. A candidate who is ignored, dominated, or threatened (however subtly) will give a poorer

performance than one who is listened to sympathetically. One of the basic faults of the testing team in "A Comedy of Errors" is their apparent disinterest in the candidate and his performance.

A good performance from the candidate is achieved by lending him psychological support. Testers should display a friendly interest in the candidate's performance. They should ask questions with a purpose and a point, and then listen attentively to the candidate's responses.

VII. The Videocassette

The videocassette contains two sample tests, "A Comedy of Errors" and the "LLC Oral Proficiency Test." The tests were originally conceived as a training film on how oral proficiency testing should and should not be conducted. The idea originated at the Language Learning Center (LLC) where "A Comedy of Errors" was first produced in black and white. The Interagency Language Roundtable's Testing Committee approved the production of a series of tapes to be made available to the government language community. Consequently, "A Comedy of Errors" was retaped in color, and the second test, the "LLC Oral Proficiency Test" showing a correct way to interview was produced.

The original LLC concept was expanded to emphasize as much as possible the commonality of technique used by testers throughout the government language training community. However, certain features are more characteristic of the LLC test than of tests conducted at FSI or DLI. For example, the four-phase division of the test is an LLC feature as is the use of two native language testers (compared to FSI's testing team comprising a native speaker and a certified language examiner). Furthermore, at FSI job-related topics are customarily discussed while the LLC test normally covers a wider range of topics.

The candidate in the sample tests is thoroughly familiar with the test, a tester himself and also a nonnative speaker of English. His English was in reality so strong that testing his real proficiency would have been less informative than testing someone at a lower level. A good actor, the candidate obliged by lowering his level, and in fact, produced speech at two different levels in the two separate tests.

The candidate's listening comprehension is high compared to his ability to speak, although it is not unusual for listening comprehension to be a + point or full point higher than speech production. The candidate uses certain colloquialisms while failing to control some grammar constructions of an even lower level, a fact often implying control of less idiomatic language at a

lower level. Finally, the number of mistakes he makes is comparatively few for the levels represented.

"A Comedy of Errors" illustrates many of the "don'ts" of oral proficiency testing. It is, of course, a parody because no real test would contain such a large collection of errors. No actual tester should take offense, because the final product is so grossly exaggerated. On the other hand, all parodies have a basis in real life. Testers sometimes fail to pay attention to the candidate. Now and then they pose a useless question. Sometimes they glance at their watches. The point is to be aware that such errors can take place, to be aware of their effect on the candidate, and to guard against them.

The videocassette should probably be viewed twice: one time for a general overview and a second time for specific details. We suggest that "A Comedy of Errors" be seen first. Doing so will sensitize the viewer to some of the basic problems in the oral interview, thus helping to stress the improved technique illustrated in the "LLC Oral Proficiency Test."

VIII. Commentaries on the Taped Tests

The following commentaries (one for each test) attempt to provide further insight into the philosophy of oral proficiency testing, elicitation technique, and rating standards. Each page of commentary is divided in half: the left side discusses the testers' questions and elicitation technique; the right side describes the candidate's answers and reactions. The commentaries are not complete transcripts of each test. Rather, they comment on selected questions and answers to provide some insight into oral interview testing methods. In "A Comedy of Errors" the candidate's speech sample is so short that the comments necessarily emphasize the testers' deportment and elicitation technique. Where the testers are remiss, the commentary suggests a suitable way to improve the testing technique.

In the "LLC Oral Proficiency Test" the emphasis shifts to a more even-handed discussion of the candidate's speech production and the testers' questions and technique. The commentary focuses on the nature of the candidate's responses and their significance for rating by citing other possible utterances that might be given under the same test conditions.

We suggest that the viewer read the commentary after seeing each test. After reading the commentary, a second viewing of the test might be useful.

Commentary to "A Comedy of Errors"

Comments on Testers	Comments on Candidate
On sitting down, Tester 1 acts both stern and disinterested. Unfortunately, this attitude pervades the whole test.	
Tester 1's first question is both too long and too esoteric. Furthermore, it begins on too high a level.	"Wait. I do not . . ." (for colloquial "I don't")
(Testers should warm the candidate up first with low-level, pleasant chatter which is unrated.)	"I don't understand."
"You don't keep up with the news?" (In the warm-up the tester should put the candidate at ease.)	
Tester 1 now discusses the candidate's performance in his presence: "We'll have to start out at a lower level!"	
"Where are you from?" (If for some reason, the test commences at too high a level, testers should make a smooth, unthreatening transition to a lower level.)	"I was born in Bogotá, Colombia." (Does the candidate control Past Tenses or is this a set phrase?)
Tester 1 proceeds to ask questions requiring only one-word answers, "yes," "no," "maybe." Three such questions appear in rapid succession. Moreover, two are on the exact same topic—a waste of time.	
Interrupting Tester 1 before he can finish his line of questioning, Tester 2 poses a question requiring a numeral. Before the candidate can respond, Tester 2 proceeds to provide him with several alternatives. (Waiting for the candidate to work it out himself provides useful information also, for it would furnish evidence of control of other numbers as well.)	Candidate's answer: "I come (for Past) in *one thousand* . . ." (implies bookish rather than colloquial knowledge of the language. Switch to the colloquial form comes only when Tester 2 feeds him the numeral.)
Tester 2's slouch displays disinterest. Covering his mouth with his hand could interfere with communication.	
Tester 2 again poses an alternative answer question, "boat, plane, car?" Testers' first five questions produced no ratable sample!	"Plane" (Candidate's response provides no ratable sample!)

Commentary to "A Comedy of Errors" (Continued)

Comments on Testers	Comments on Candidate
Tester 2: "How long a trip was it?"	"It was more or less . . . seven . . . seven hours from Bogotá to Miami . . ." (Last word mispronounced.)
Tester 2: "Tell me about Bogotá."	Answers rather simplistic: "Well B . . . It is a city. It is big . . . and . . . eh . . . there are many, many peoples (wrong plural). The city is very hot . . . eh . . . It is nice!" (Short sentences, usually grammatically correct. Indicative of Level 1. A higher level answer would combine short sentences into longer syntactic units and/or handle higher level topics with higher level details--which the candidate fails to do.
Tester 1 slouches practically falling asleep, yet still managing to tap his fingers impatiently on the side of his chair.	
Tester 2 queries, "What do you *know* about America?" The question's focus is too broad and the phrasing implies that the candidate did not say enough about Bogotá.	The candidate is compelled to seek clarification. "You mean the United States?"
Tester 1 interrupts the candidate before he can formulate a response. Obviously, the candidate does not yet know the North American usage of 'America' and thus provides the testers with an indication of his knowledge of 'American English'. (The testers might probe further. "Is there a difference between 'America' and 'North America'?") On the other hand, having obtained this information, the testers show insensitivity in not making a smooth transition away from the problem.	
Tester 1's gesticulations and insistent tone can make the candidate tense.	
Tester 2: "Tell us something about American holidays." Again, the question's focus is too broad. (A possible transition: "Now that you've been to America, could you pick out your favorite American holiday?" If the candidate answers	"American holidays?"

Commentary to "A Comedy of Errors" (Continued)

Comments on Testers	Comments on Candidate
positively, the tester might proceed, "How would you tell someone back home about it?")	
Tester 1 attempts to save the topic: "Thanksgiving, maybe?" And then he remembers to check with the other tester, "Would that be all right?" (Planning for this question and tester-tester as well as tester-candidate interaction all leave something to be desired.)	
Tester 2 slouches back again.	
Testers 1 & 2 ignore the candidate and begin a personal conversation. They show their first real interest in the test, but unfortunately it is not in the candidate. (A planned conversation between testers is an excellent listening comprehension check or way of setting the stage for bringing the candidate into the conversation. The testers have failed to do this because America's mania for Thanksgiving football is lost on the candidate, a foreigner.)	
Football to a non-American is a cultural problem in another sense, for "their football" is "our soccer." Until the problem appears, the testers seem oblivious to it. And when it surfaces, they are brusque in their treatment of the candidate's attempt at clarification.	"American football?"
Tester 1: "I think he means soccer!" "Why don't you say what you mean?"—does nothing to clarify the situation.	"I'm confused."
Tester insults candidate rather than encouraging him to continue his line of thought.	
Tester 2: "Where did you learn English?" This question, which has the advantage of permitting the candidate to talk on a known subject, also might mislead testers into rating on previous training, not on the	"I go to school in Bogotá." (Present Tense for Past "I went.") Control of "too" and "also" weak.

Commentary to "A Comedy of Errors" (Continued)

Comments on Testers	Comments on Candidate
sample produced in the test. For the latter reason, such questions are usually avoided at FSI. In the present test, the question has the unfortunate effect of leading the testers into discussing the candidate's previous teachers derogatorily. All such discussion should be avoided during the test and in the presence of the candidate.	
Tester 2: "What can you say about sports?" Again, the question's focus is too broad. Phrasing may imply that the candidate has not said much about anything else.	"I think sports are . . . eh . . . good for you!" (Answer is simplistic, and if the candidate were allowed to continue, he might well not add significantly to this rather straightforward statement. Contrast this with a higher-level version, "Sports are good for you since they excercise the lungs, keep you in shape and provide variation in an otherwise dull routine!")
Tester 1 abruptly shifts topic without permitting candidate to complete his answer. This action deprives testers of a ratable sample.	
Tester 2: "Do you have any other interests?" The tone in which the question is asked is problematical. The phrasing and the tone may imply that the candidate couldn't talk much about previous topic, perhaps tester can find one for him that he can perform better on. The tone also implies that they doubt it.	
Testers ignore several excellent topics: 'movies,' 'women.' The first is well suited as a probe for Level 2 description and narration, and the second, 'women,' always results in some speech production.	'womens' (Wrong plural!)
Tester 2 shifts position in chair, bored stiff.	
In an insistent tone, Tester #1 picks up on reading. "What kinds of things do you read?" Then fails to follow up on 'newspapers' a good Level 2 or Level 3 probe for description, narration and indirect discourse structures.	"I read the newspaper."

Commentary to "A Comedy of Errors" (Continued)

Comments on Testers	Comments on Candidate
Tester 1: "What about Shakespeare?" The topic is at too high a level for the candidate and could not possibly be even a probe. Also, the material is too academic for most tests even if the level were suitable.	"I know Shakespeare." (for "I know *about* Shakespeare.")
Tester 2 interrupts his colleague to extend the question. "Since you have studied English, I *do assume* you know *something* about Shakespeare!" Asked in a tone implying that everybody has heard of the Bard of Avon.	"I don't comprehend the question." (Wrong verb, should be "understand.")
Tester 2 fails to perceive that the questions are clearly above the candidate's level. "Who would you pick for the lead in *Henry the Fourth?*" The question assumes detailed knowledge of English culture, Shakespeare, the play itself, the qualities of the play's characters and of English-speaking actors. The question tests factual, not linguistic, content.	"I cannot answer." (Object missing for "I cannot answer that/the question.")
Tester 2: "That's probably just about enough" in a tone which implies that the candidate did not do very well.	
Testers discuss the candidate's performance in front of him.	
The candidate is dismissed in the same abrupt fashion in which the whole test has been conducted.	

"THE BITTER END"

BASIC QUESTIONS: Did the testers procure a sample adequate enough to accurately rate the candidate's speech production?

Were there enough listening comprehension probes to adequately rate the candidate's understanding in the target language?

How could the testers have conducted the interview so that the candidate would have produced more and they would have said less?

SCORING: The speech sample was much too short to yield an accurate rating of the examinee's proficiency. But short as it is, the candidate's performance hovers around Level 1, although a good testing team would insist on probing more extensively both for breadth at Level 1 and for higher-level abilities at Levels 2 & 3.

Commentary to "LLC Oral Proficiency Test"

Comments on Testers	Comments on Candidate
WARM-UP (This phase of the test puts the candidate at ease, reacquaints him with the language if need be, and gives the testers a preliminary indication of the candidate's level of proficiency.)	Note the informal arrangement of the testing room with comfortable chairs and pleasant decor to relax the candidate.
Tester 1: "Where are you from?" (This question permits candidate to answer either in the Present Tense or in the Past.)	"I was born in Bogotá, Colombia." (Does the candidate control Past Tenses or is this a set phrase?)
Tester 1: "And when did you come here to this country?" (Having heard that the candidate controls at least one Past Tense, the tester proceeds to check if he has others. The general topic of the conversation also permits developing it further in the Past.)	"I came in 1972." (Note the Past Tense of a slightly less frequent verb. Note also the normal form of the numeral 'nineteen hundred' vs. the bookish 'one thousand,' etc., of "A Comedy of Errors.")
Tester 1: "How long have you been here now?" (This question could test candidate's control of the Present Perfect Tense extending into present time, if the candidate had given a longer answer. A clear indicator of poor control would have been, "I am here for three years.")	"Three years." (Answers are succinct, but grammatically accurate and highly appropriate suggesting that the candidate is beyond the bookish stage and controls the language in a somewhat colloquial fashion.)
Tester 1: "And where did you study English?" (Tester's tone is objective, not threatening as in "A Comedy of Errors.")	"I went to school in Bogotá . . . more or less, two years." (Control of an irregular Past Tense form. Insertion of "more or less?" here not English. Perhaps, "almost" would be more idiomatic.)
Tester 1: "What kind of school do you go to here?" (A possible rephrasing with a higher-level vocabulary item would include "attend.")	"I want to learn to speak Spanish." (A common slip of the tongue when the candidate has to concentrate on how to speak in the target language. He corrects it to *English* immediately. This error is in information so would not count against him, especially since he corrected it.)
(Warm-Up questions are often couched in simple terms, that is, Level 1 language. The candidate has shown that he can function at Level 1, so the testers proceed to check if he can function at a higher level. Their tentative hypothesis places him on Level 2 which they now test in the next phase of the interview.)	

Commentary to "LLC Oral Proficiency Test" (Continued)

Comments on Testers	Comments on Candidate
LEVEL CHECK (SPEAKING: This phase attempts to find the candidate's level of speaking proficiency and to test the breadth of grammatical structures, vocabulary, and idioms at that level. LISTENING COMPREHENSION: This phase tests for the candidate's level in understanding of the questions and discussion.)	"I go to a school . . ." (inappropriate to this context; either add a suitable location, e.g., in the District, etc., or "I'm going to school.")
Tester 2: "Where do you live now?"	"I live in a house now. I live in Virginia." (Sentences complete, simple and grammatically accurate. However, he did not combine them in a Level 2 fashion, e.g., "I live in a house located in Virgina.")
PROBE (This phase of the test is often intermingled with Phase II, the Level Check. SPEAKING: Probes are employed to find the level at which the candidate can no longer perform. LISTENING COMPREHENSION: Probes are used to find the level at which the candidate can no longer understand.) Tester 2: "What do you see are the advantages of living in a house over an apartment?" (This probe is an attempt to shift the level higher. Its success depends on the level of the candidate's response.)	"I think that a house is better . . . is better when you have children." (Despite this higher-level sentence, the candidate otherwise sticks to fairly simple sentences. The candidate barely shifted levels at all.)
	"The children can play. They can make noise. It is a problem."
Tester 2: "Is your house outside the city in the suburbs?" (A better phrasing for listening comprehension would choose one of the built-in alternatives: *outside the city* or *in the suburbs*, and use only the one phrase in the question.)	
Tester 2: "Describe for me your daily routine." (Description is normally regarded as a Level 2 task. A common topic at Level 2 is one's own life. Hence, this question. However, it is not the fact that the candidate discusses the topic, but *how* he discusses it that is the determining factor in rating his speech production.)	"I waked up." (for "I woke up.") (Control of Past Tense has been solid up to now. Verb is not as common as "come" and "be." This error could easily occur as the isolated mistake of a speaker at an even higher level.)

Commentary to "LLC Oral Proficiency Test" (Continued)

Comments on Testers	Comments on Candidate
	"I took a shower. I shaved. I dressed myself." (Not usually reflexive in English. Cf. colloquial "I got dressed.")
	"I ate . . ." (Starts to drop topic or grope for an answer.)
Tester 2: Unwilling to let a possible ratable sample go, the tester tries to convert a painful pause into a productive one. "What did you eat for breakfast?"	
Tester 2: "Could you describe how you got to work?"	"I go to school." (Usually, an English speaker in this context would say either "I go to school" plus a time or location phrase or "I'm going to school now.")
	"I'm not working now."
	"I took my car and drove to school."
	The candidate controls English simple Past Tense well. He can narrate (which includes description).
Tester 2: "It's rather an experience to get back and forth. Do you have a car pool?"	Candidate asks for clarification. "You mean, if I'm with somebody?" (Phrasing here odd from an English point of view. More natural: "Do you mean, do I drive, go with someone to work?") "No, I don't."
Tester 1: "What about the overall cost of living?" (Testing for comprehension of a phrase well-known to any attentive listener of radio broadcasts, viewer of television news or reader of newspapers or general magazines.)	"What do you mean?" (Phrase stumps candidate. When he finally understands, he winds up with a simplified view without details. "I think life is more expensive.")
Tester 1 rephrases, "How do you find . . .?"	
"What do you think causes it to be more expensive?" (Having introduced the topic, the testers want to get the most ratable sample out of it and so pursue it by asking for the candidate's opinion. "I agree," "I disagree," etc.)	"Well, Latin American countries are not more . . . (gropes) . . ."

Commentary to "LLC Oral Proficiency Test" (Continued)

Comments on Testers	Comments on Candidate
Supported opinion is generally indicative of mid and high levels. The candidate needs to produce longer sentences, subordinate clauses, etc. This the candidate fails to do.	
Tester 2: It is not normal policy for testers to provide answers to the candidate (see the problem this led to in "A Comedy of Errors"). However, if the candidate keeps mulling over what the right answer might be and the testers wish to remain with the topic, then it is easier to provide the answer and go on. Here, "less developed."	The candidate searches for a word, finally resorts to a catchall phrase, "many things."
Tester 2 provides a smooth transition to the next topic. "I'd like to switch the topic if I might."	
Tester 2 tries a hypothetical question and, finding that the candidate fails to grasp it, rephrases to something easier. Original: "You have responsibility for determining domestic policy, what problems would you face?" Paraphrase: "If you were President, what problems do you think you would have to solve? What are the major problems before you?"	Candidate asks, "Can you repeat that? I don't think I understand."
	Candidate struggles . . . "If I was . . . I'm going to . . ." (The candidate has difficulty with the English conditional forms, sequence of tenses, etc. Probably not higher than Level 2+ if this pattern of errors continues.)
Tester 2 breaks the larger question down into shorter sentences to test the candidate's listening comprehension in vocabulary.	"I know, what I want (leaves out "to say"), but I don't have the words . . . eh . . . to say . . . you know . . . what I'm going to do." (Candidate understands, but lacks the vocabulary to deal with the problem. The candidate is clearly not Level 3 on this topic, for it is assumed that Level 3 controls all the basic vocabulary and can use circumlocutions to make himself understood for difficult vocabulary items he does not know. Candidate does not perform like a Level 3 speaker.)
	"The young people . . . have a problem with . . . eh . . . with . . . eh . . . *the* drugs." (Article not used in English.)

Commentary to "LLC ORAL Proficiency Test" (Continued)

Comments on Testers	Comments on Candidate
	"I would think about *to* control . . . " (Candidate fails to nominalize *control* to *controlling*. Infinitive marker out of place here.)
	Sums up by simplification. "Those are big problems." Fails to deal with their complexity in a linguistically sophisticated sense. Therefore, not at a high level.
Tester 1 provides a good lead into the question to check interest.	"You mean what I do?"
PROBE: Tester 1: "I wonder if you could tell us . . . You've been here. I think you've said you've been here three years. I wonder if you could tell us about Thanksgiving." (Used as a probe for Future Tense forms.)	Candidate avoids facts: the holiday's origin, history of its official declaration, etc. He stays with what he can talk about.
The Tester is flexible, permitting the candidate to answer in the area in which he is most comfortable.	"My sister is coming to the *area of Washington*." (Not idiomatic. Cf. Washington Area. Candidate's answer is a typical nonnative formation.)
	"She's going to spend *the* time with us." (*The* not suitable in this context.)
	"I'm going to be with my sister . . . *more or less*." (The "more or less" here strikes American English ears as odd. It is not necessarily wrong, but it is definitely out of place in an idiomatically correct sentence for this topic.)
WIND-UP: (This phase returns the candidate to the level at which he performed best. It is designed to leave the candidate with a feeling of accomplishment after thorough probing. To do this, testers choose topics of conversation on a slightly lower level more comfortable for the candidate.)	
Tester 2: "What do you usually do in your free time?"	"I play with the children."

Commentary to "LLC ORAL Proficiency Test" (Continued)

Comments on Testers

Tester 2: "Do you ever get to go to the movies?"

Comments on Candidate

"My wife *wants* the movies." (Corrects to *like*, finally *likes*. This shows some problem in control of verb form and congruence. A more idiomatic sentence would omit *the*.)

"I like the movies *toos*." (Corrects to *too*.)

Testers wind up test in a friendly manner.

BASIC QUESTIONS: Did the testers procure a sample adequate enough to accurately rate the candidate's speech production?

Were there enough listening comprehension probes to adequately rate the candidate's understanding in the target language?

What were, in your opinion, the crucial differences between the present test and "A Comedy of Errors?"

SCORING: Testers base a rating on a pattern of errors, never on a single one. A single error counts nothing; it must be confirmed by other errors which reveal the candidate's inability to function at a particular level. Thus, for example, the candidate's use of "waked up" instead of "woke up" has no significance in itself. If there had been other wrong Past Tense forms, this error would be meaningful.

The following comments on the candidate's speech sample summarize his errors and provide the basis for the final rating.

PRONUNCIATION: Slightly, but not markedly, foreign.

FLUENCY: Still somewhat hesitant, but with more "productive" pauses than "painful" ones.

VOCABULARY: Somewhat limited if considering candidate for Level 3. Furthermore, candidate could not use circumlocution (a Level 3 task).

COMPREHENSION: Clearly better than speech production, but without being able to understand to a high degree more complex formal discourse.

GRAMMAR: a. Candidate controls all but the most irregular Past Tenses (cf. "waked up").

b. Problems with proper use of Progressive Tense. (This tense is so idiomatic that full control of it may still elude a good Level 4 speaker.)

c. Trouble with English Conditionals (solid control of Conditionals indicates Level 3).

Commentary to "LLC Oral Proficiency Test" (Continued)

d. Most simple and short combined sentences grammatically correct. Longer combined sentences handled with hesitancy, sometimes incomplete and often grammatically wrong.

e. Candidate can describe and narrate to some extent (indicative of Level 2).

f. Problems with nominalizations: "think to control" for "think about controlling." (Nominalization is a higher-level skill. Again, evidence that the candidate is not in the higher range.)

g. Articles, nominalizations, prepositions and Progressive Tense problems may be indicative of several levels depending on the extent of errors. Sporadic errors occur as late as Level 4/4+, while more frequent errors occur lower on the scale.

OVERALL SPEAKING RATING: 2

APPENDIX

Detailed Description of the FSI Checklist

The main purpose of the Checklist of Performance Factors is to provide a profile of the principal elements in an examinee's spoken language proficiency. In addition, the Checklist can be a useful guide in determining the S-Rating that should be assigned.

Immediately after the speaking test is over, a mark should be made on each of the five scales of the Checklist. The marks should be placed in accordance with the descriptions of the blanks on each scale given below. If the examinee's performance on any factor falls between two blanks, the scale should be marked accordingly and scored half-way between.

In these descriptions, 1 denotes the left-most blank in each scale and 6 denotes the right-most blank.

Accent

1. Pronunciation frequently unintelligible.
2. Frequent gross errors and a very heavy accent make understanding difficult, require frequent repetition.
3. "Foreign accent" requires concentrated listening and mispronunciations lead to occasional misunderstanding and apparent errors in grammar or vocabulary.
4. Marked "foreign accent" and occasional mispronunciations which do not interfere with understanding.
5. No conspicuous mispronunciations, but would not be taken for a native speaker.
6. Native pronunciation, with no trace of "foreign accent."

Grammar

1. Grammar almost entirely inaccurate except in stock phrases.
2. Constant errors showing control of very few major patterns and frequently preventing communication.
3. Frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding.
4. Occasional errors showing imperfect control of some patterns but no weakness that causes misunderstanding.
5. Few errors, with no patterns of failure.
6. No more than two errors during the interview.

Vocabulary

1. Vocabulary inadequate for even the simplest conversation.
2. Vocabulary limited to basic personal and survival areas (time, food, transportation, family, etc.).
3. Choice of words sometimes inaccurate, limitations of vocabulary prevent discussion of some common professional and social topics.
4. Professional vocabulary adequate to discuss special interests; general vocabulary permits discussion of any nontechnical subject with some circumlocutions.
5. Professional vocabulary broad and precise; general vocabulary adequate to cope with complex practical problems and varied social situations.
6. Vocabulary apparently as accurate and extensive as that of an educated native speaker.

Fluency

1. Speech is so halting and fragmentary that conversation is virtually impossible.
2. Speech is very slow and uneven except for short or routine sentences.
3. Speech is frequently hesitant and jerky; sentences may be left uncompleted.
4. Speech is occasionally hesitant, with some unevenness caused by rephrasing and groping for words.
5. Speech is effortless and smooth, but perceptibly nonnative in speed and evenness.
6. Speech on all profesional and general topics as effortless and smooth as a native speaker's.

Comprehension

1. Understands too little for the simplest type of conversation.
2. Understands only slow, very simple speech on common social and touristic topics; requires constant repetition and rephrasing.
3. Understands careful, somewhat simplified speech directed to him, with considerable repetition and rephrasing.
4. Understands quite well normal educated speech directed to him, but requires occasional repetition or rephrasing.
5. Understands everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech.
6. Understands everything in both formal and colloquial speech, to be expected of an educated native speaker.

Instructions for Use of the Checklist to Determine S-Ratings

	1	2	3	4	5	6	(A)
Accent	4	3	2	2	1	0	_____
Grammar	6	12	18	24	30	36	_____
Vocabulary	4	8	12	16	20	24	_____
Fluency	2	4	6	8	10	12	_____
Comprehension	4	8	12	15	19	23	_____
						Total	_____

Procedure: Place in Column (A) the credits to be given for each scale on the *Checklist*. For example, a check mark in position 3 on the "Accent" scale is given a credit of 2. Add the credits to find the total score. The final S-Rating is to be equated with the total score by the following table.

Score	Rating	Score	Rating	Score	Rating
16-25	S-0+	43-52	S-2	73-82	S-3+
26-32	S-1	53-62	S-2+	83-92	S-4
33-42	S-1+	63-72	S-3	93-99	S-4+

If the total scores by different examiners yield different S-Ratings, an average of the total scores should be used to determine the final rating.

Instructions for the Use of the Checklist

The purpose of the Checklist is to provide a record of your evaluation of certain aspects of an individual's spoken language proficiency.

On the Checklist you will find listed a series of factors: Accent, Grammar, Vocabulary, Fluency, and Comprehension. You are to rate the individual's performance by placing a check mark on the scale beside each factor.

We will use the factor "Accent" as an example of how to use the scales.

If you feel that an individual's accent is very closely related to one end of the scale (i.e., his accent is *extremely* foreign, or *extremely* native), place a check mark as follows:

Foreign ___X___ : ___ : ___ : ___ : ___ Native

or

Foreign ___ : ___ : ___ : ___ : ___X___ Native

If you feel that his accent is quite closely related to one end of the scale or the other (i.e., is *quite* foreign, or *quite* native), place your check mark as follows:

Foreign ___ :X: ___ : ___ : ___ Native

or

Foreign ___ : ___ : ___ : X : ___ Native

If you feel that his accent is only slightly related to one end of the scale as opposed to the other (i.e., is *more foreign than native*, or *more native than foreign*), then place your check mark as follows:

Foreign ___ : ___ :X: ___ : ___ Native

or

Foreign ___ : ___ : ___ :X: ___ Native

As an aid to accuracy and uniformity in rating, each of the factors is explained below:

1. *Accent*

   Consider pronunciation of individual sounds (consonants and vowels) as well as pitch and stress.

2. *Grammar*

   Consider individual words and arrangements of words for grammatical accuracy.

3. *Vocabulary*

   Consider adequacy of vocabulary (both sufficiency and accuracy) for the topics discussed.

4. *Fluency*

   Consider general smoothness or evenness in performance.

5. *Comprehension*

   Consider both completeness and speed of understanding.

These factors are not necessarily interrelated; therefore, you should not expect that an individual will necessarily receive the same rating on all the factors.

Consider each factor separately and in the order listed.

Bibliography on the Oral Interview

Clark, John L. D. 1972. *Foreign Language Testing: Theory & Practice.* Philadelphia: Center for Curriculum Development. Pp. 121-129 on the "FSI Language Proficiency Interview." (Available from Didier, 29 Lexington Road, Concord, MA.)

Defense Language Institute. English Language Center. 1975. *Handbook for Conducting Task Analyses and Developing Criterion-Referenced Tests of Language Skills.* [San Antonio.] March 1975.

Jones, Randall L. 1975. "Testing Speaking Proficiency" in *Testing Language Proficiency,* ed. Randall L. Jones & Bernard Spolsky, Arlington, Va: Center for Applied Linguistics. Pp. 1-7.

———, (To Appear). "Achieving Objectivity in Subjective Language Tests" in *The Proceedings of the Fourth International Congress of Applied Linguistics.* Stuttgart.

Lowe, Pardee, Jr. 1975. "The Oral Interview—A Criterion-Referenced Test." Paper delivered in the Section on "Foreign Language Testing: A Time for New Directions" at the 1975 FIPLV/AATG/ACTFL Convention in Washington, DC, in November 1975. (mimeo). 15 pp. Available ERIC ED 119 473.

Weinstein, Allen I. 1975. "Foreign Language Majors: The Washington Perspective." *ADFL Bulletin.* (6 May 1975): pp. 18-27.

Wilds, Claudia P. 1975. "The Oral Interview Test" in *Testing Language Proficiency,* ed. Randall L. Jones & Bernard Spolsky, Arlington, Va: Center for Applied Linguistics. Pp. 29-38.